Pitch strength of normal and dysphonic voices

Rahul ShrivastavDavid A. EddinsSupraja AnandWPS

Citation: J. Acoust. Soc. Am. **131**, (2012); doi: 10.1121/1.3681937
View online: http://dx.doi.org/10.1121/1.3681937
View Table of Contents: http://asa.scitation.org/toc/jas/131/3
Published by the Acoustical Society of America

Pitch strength of normal and dysphonic voices

Rahul Shrivastav[a)]
Malcom Randall VAMC and University of Florida, Dauer Hall, P.O. Box 117420, Gainesville, Florida 32611

David A. Eddins
University of South Florida, 4202 Fowler Avenue, PCD 1017, Tampa, Florida 32620

Supraja Anand
University of Florida, Dauer Hall, PO Box 117420, Gainesville, Florida 32611

(Received 26 April 2011; revised 14 December 2011; accepted 2 January 2012)

Two sounds with the same pitch may vary from each other based on saliency of their pitch sensation. This perceptual attribute is called "pitch strength." The study of voice pitch strength may be important in quantifying of normal and pathological qualities. The present study investigated how pitch strength varies across normal and dysphonic voices. A set of voices (vowel /a/) selected from the Kay Elemetrics Disordered Voice Database served as the stimuli. These stimuli demonstrated a wide range of voice quality. Ten listeners judged the pitch strength of these stimuli in an anchored magnitude estimation task. On a given trial, listeners heard three different stimuli. The first stimulus represented very low pitch strength (wide-band noise), the second stimulus consisted of the target voice and the third stimulus represented very high pitch strength (pure tone). Listeners estimated pitch strength of the target voice by positioning a continuous slider labeled with values between 0 and 1, reflecting the two anchor stimuli. Results revealed that listeners can judge pitch strength reliably in dysphonic voices. Moderate to high correlations with perceptual judgments of voice quality suggest that pitch strength may contribute to voice quality judgments. © *2012 Acoustical Society of America*. [DOI: 10.1121/1.3681937]

PACS number(s): 43.71.Bp, 43.71.Gv, 43.72.Ar [WPS] Pages: 2261–2269

I. INTRODUCTION

Voiced speech stimuli are typically described as having three perceptual attributes—pitch, loudness, and quality. Pitch is defined as "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high" (ANSI, 1994). For voices and non-vocal complex tones, perceived pitch is related to complex interactions among the stimulus harmonic structure and details of the magnitude and phase spectra as well as characteristics of the auditory system (e.g., see Moore *et al.*, 1997). Variation in the pitch of running speech carries prosodic information, while static differences in pitch may contribute to talker identification. Two sounds that are perceived to have the same pitch may differ in terms of the prominence or saliency of the pitch sensation that they evoke. For example, when the same note is produced by two musical instruments, such as stringed (e.g., guitar) and wind instruments (e.g., flute), the note produced from a stringed instrument typically results in the perception of a more prominent pitch than that of a wind instrument. Likewise, a 500 Hz pure tone and a bandpass filtered noise centered on 500 Hz are perceived to have the same pitch, but the band-pass filtered noise evokes a weaker pitch sensation. This perceptual attribute is called the pitch strength of the sound and is independent from pitch itself.

The pitch strength of a sound is affected by a number of changes to the acoustic signal Zwicker and Fastl (1990) report the pitch strength of a variety of stimuli in three different frequency regions (125, 250, and 500 Hz). These included pure tones, complex tones, amplitude-modulated tones, narrow-band noise, broad-band noise, band-pass noise, and comb filtered noise. The results indicated that pure tones evoke the greatest pitch strength followed by complex tones and noise stimuli. Various noise stimuli elicited pitch strength values that were smaller by a factor of 5 or 10 relative to the pure tone stimuli of equal pitch. These results are consistent with the notion that pitch strength varies on the continuum of periodic vs stochastic stimuli. Furthermore, Zwicker and Fastl (1990) observed that certain aspects of the noise stimuli, such as the cut-off frequency and spectral slope, also affected their pitch strength. High-pass filtered noise with lower cut-off frequencies produced very low pitch strength relative to the different types of tonal stimuli. Pitch strength of a stimulus also increased as the steepness of the spectral/filter slope increased.

Psychoacoustic experiments to evaluate pitch and pitch strength often require listeners to match the pitch strength of a test signal to that of "iterated rippled noise" (IRN). IRN is a class of stimuli generated by attenuating and adding a delayed version of a broad-band noise to itself, such that the stimulus has regularly-spaced spectral peaks that resemble a harmonic tonal complex with a relatively flat temporal envelope lacking obvious envelope periodicity (Fastl and Stoll, 1979; Fastl, 1988; Leek and Summers, 2001; Patterson *et al.*, 1996; Shofner and Selas, 2002; Yost *et al.*, 1978; 1979; 1994; 1996; Yost, 1982; 1996; 1997). The pitch strength of an IRN can be systematically varied through further modifications of the

a)Author to whom correspondence should be addressed. Electronic mail: rahul@msu.edu

parameters used to create the IRN. For example, varying the delay duration (d, in ms), the level of attenuation of each iteration of the noise (a, in dB), and/or increasing the number of iterations (n) itself can result in systematic variations in pitch strength (Yost 1996; Yost *et al.*, 1996; Patterson *et al.*, 1996). An example IRN circuit is shown in Fig. 1. As n increases, the tonal component of the perception grows stronger. With increasing attenuation (a), the IRN stimulus resembles the original broad-band noise more closely, and evokes a faint pitch sensation, i.e., lower pitch strength. Yost and his colleagues (Yost *et al.*, 1978; 1994; 1996) as well as Patterson *et al.* (1996) have demonstrated that the pitch strength of an IRN stimulus is proportional to the height of the first autocorrelation peak of the IRN waveform. Informal observation indicates that pitch strength varies in speech as well. Certain speech sounds, such as vowels, are highly periodic and elicit a strong pitch sensation. In contrast, other sounds like fricative consonants may elicit a weak pitch sensation. Many of the acoustic changes observed to affect the pitch strength of complex tones and noise stimuli are also commonly observed in speech. For example, factors such as spectral slope or the relative noise levels are frequently observed to change within and across speakers, and are often correlated with changes in voice quality (Klatt and Klatt, 1990; Shrivastav and Sapienza, 2003). Therefore, it is possible that pitch strength and certain aspects of perceived voice quality are related percepts. While research on pitch strength typically focuses on noise (Fastl and Stoll, 1979; Leek and Summers, 2001; Patterson *et al.*, 1996; Shofner and Selas, 2002; Yost *et al.*, 1978; 1979; 1994; 1996; Yost, 1982; 1996; 1997) or relatively simple harmonic sounds (Fastl and Stoll, 1979; Shofner and Selas, 2002), to our knowledge there are no empirical studies examining the pitch strength of a voice or how that pitch strength may affect judgments of its quality. In the present study, the anchored magnitude estimation task of Shofner and Selas (2002) is adapted to estimate the pitch strength associated with voice samples differing along the voice quality dimensions of breathiness and roughness.

It is important to distinguish pitch strength from pitch itself, as well as from descriptors such as "voice quality" and "timbre." In terms of vocal quality, singers often use terms like "rich," "dry," "bright," or "flat" to describe a voice, whereas speech pathologists and voice scientists describe vocal qualities using terms such as "breathy," "rough," "strained," etc. (ASHA 2002; Colton and Casper, 1996). The voice samples chosen for the current study vary along accepted dimensions of voice quality with each dimension encompassing a continuum of voices ranging from normal quality to severely disordered or "dysphonic" quality. Dysphonic voice quality may be defined as a voice quality that is not appropriate for the age, sex, gender, or culture of the talker. Such descriptions are qualitative in nature, although terms used to describe dysphonia are often explained in the context of vocal fold physiology and/or specific acoustic characteristics of the vocal signals.

The voice qualities of roughness and breathiness are two of several commonly studied voice quality percepts (Kreiman and Garrett, 2000) and are characteristic of most voices. Breathiness and roughness are particularly noteworthy in the context of disordered voices, since voice quality is often used as an indicator of voice pathology (Kent, 1996). Specifically, vocal breathiness may be defined as audible air escape in the voice (Kempster *et al.*, 2009). Vocal roughness may be defined as the perceived irregularity in the voicing source (Kempster *et al.*, 2009). Furthermore, these two qualities are not mutually independent. It is frequently the case that roughness and breathiness co-occur in dysphonia. Because of their clinical relevance and correlations to numerous physical and neurological pathologies, understanding potential acoustic, perceptual, and physiological correlates to voice quality percepts is an essential component of voice research and clinical practice.

Timbre has been defined as "that attribute of auditory sensation which enables a listener to judge that two nonidentical sounds, similarly presented and having the same loudness and pitch, are dissimilar" (ANSI, 1994). Such a definition is rather limited in scope, as sounds that differ in pitch and loudness may also differ in timbre. Nevertheless, the most dominant acoustic attribute contributing to timbre differences is overall spectral shape (Houtsma, 1997). In terms of speech, as noted by Houtsma (1997), the pitch contour associated with vowels and voiced consonants is related to quasi periodic vocal fold vibrations, which in turn is partly characteristic of a given talker. In contrast, robust differences in spectral shape give rise to different vowels of a language and are distinguished perceptually by timbral differences. Within a phonemic category, timbre may also differ substantially within and across talkers. Houtsma (1997) speculated that vocal pitch and timbre are largely independently of each other, based on the assumption that, to a first approximation, vocal fold vibration, and vocal track resonances are not strongly dependent. To minimize timbre differences across talkers, the natural voice samples studied here are restricted to the single phoneme /a/, as in the American English word "hot." Nevertheless, the sustained voiced samples from different talkers varied in fundamental frequency, temporal characteristics, and spectral shape.

The goals of the present study were to (i) determine if listeners can judge reliably the pitch strength of voices selected along the continuum of normal to severely dysphonic breathy and rough voice quality and (ii) to determine the relation, if any, between pitch strength and vocal breathiness and roughness. The long-term goals of this work are to develop better and more accurate methods to characterize dysphonic speech

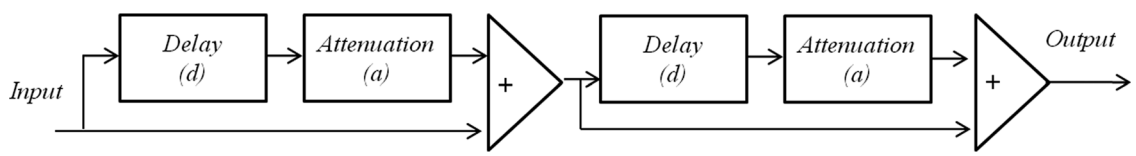

FIG. 1. A schematic diagram of the circuit used to generate IRN stimuli with specific delay (d) and attenuation (a) parameters.

in both laboratory and clinical settings, to improve our overall understanding of dysphonia and its perceptual and acoustic analogues, and to develop and improve clinical tools that positively impact patients with dysphonia.

II. METHOD

A. Listeners

Ten female graduate students from Department of Speech-Language and Hearing Sciences at the University of Florida were recruited for the study.[1] The age of the listeners averaged 22 years and ranged from 20 to 25 years. All listeners were native speakers of American English and passed a hearing screening at 20 dB HL at octave frequencies between 250 and 4000 Hz (ANSI, 2004). Listeners were compensated monetarily for participating in this study. Although these listeners were familiar with dysphonic voice qualities through their academic coursework, they had no prior experience in judging pitch strength. All procedures were approved by the Institutional Review Board at the University of Florida and all listeners voluntarily consented to participation.

B. Instrumentation

All experimental procedures were controlled through the TDT System III hardware and software (Tucker-Davis Technologies, Inc., Alachua, FL). The hardware included the RP2 DSP and D/A module, programmable attenuators (PA5), a headphone preamplifier (HB7) and Etymotic ER2 insert ear transducers (Etymotic Research, Inc., Elk Grove Village, IL). The stimulus presentation and data acquisition was controlled using the SYKOFIZX software application (Tucker-Davis Technologies, Inc, Alachua, FL). All listening sessions were conducted in a single-walled sound booth and the stimuli were delivered at 85 dB SPL in the right ear.

C. Procedures

Previous research (Fastl and Stoll, 1979; Shofner and Selas, 2002) has shown that pitch strength can be scaled using direct magnitude estimation to obtain listener judgments. Therefore, a magnitude estimation task with anchor stimuli as described by Shofner and Selas (2002) was adapted in this experiment. Listeners heard three different stimuli on each trial, separated by 500 ms of silence. The first item was an anchor with very low pitch strength (wideband noise) and was assigned a pitch strength value of 0. The second item consisted of the test stimulus and was assigned a pitch strength value by the listener. The third item was an anchor with very high pitch strength (1000 Hz pure tone) and was assigned a pitch strength value of 1.[2] Listeners were asked to judge the pitch strength of the test stimulus on each trial by positioning a continuous slider between the values of 0 and 1. The distance between the two anchors was calibrated into 100 equidistant steps. Therefore, listener judgments could range in values from 0 to 100.

1. Pitch strength of IRN (training)

Since the listeners tested in this experiment had no previous experience in judging pitch strength, a training task was developed which mirrored the main experiment. Listeners were asked to judge the pitch strength of the five IRN training stimuli using the anchored magnitude estimation task described above. Each stimulus was presented 10 times in random order, resulting in a total of 50 items for each listener (5 levels of attenuation $\times$ 10 repetitions). The data from the 10 repetitions of the stimulus were averaged to obtain a single score for each attenuation level of the IRN stimuli. The training task took approximately 20 minutes for each listener.

2. Pitch strength of vowels (experiment)

Listeners judged the pitch strength of the 21 dysphonic voices with the same anchored magnitude estimation task as used in the training paradigm. Each voice was judged ten times resulting in a total of 210 stimuli (21 voices $\times$ 10 repetitions). The order of presentation of these stimuli was randomized. Listeners were tested in a single test session which lasted for two hours. However, a short break was provided approximately every 10 minutes to minimize fatigue.

D. Stimuli

Two sets of stimuli were created for this experiment. The first set consisted of IRN stimuli, and was used for training listeners to judge pitch strength. The second set of stimuli consisted of dysphonic voices and was used for the main experiment.

1. Training stimuli

A set of five IRN stimuli was created for the training task. For each stimulus, a broadband noise was generated and lowpass filtered at 10 000 Hz. To this noise was added a delayed and attenuated copy of itself, creating a single iteration. The final IRN was created with 10 iterations using a fixed delay of 16 ms and five attenuation values ranging from 0 to 16 dB in steps of 4 dB. A delay of 16 ms corresponds to a fundamental frequency of 62.5 Hz, and was chosen to be outside the range fundamental frequencies of the test stimuli (67 to 257 Hz). These were selected as training stimuli because prior research has shown these to vary systematically in their pitch strength (Yost, 1996; Shofner and Selas, 2002) with higher attenuation resulting in the lower pitch strength. Attenuation level of 0 dB represented "high pitch strength" and attenuation level of 16 dB resulted in stimuli with the least pitch strength. The duration of each stimulus was 500 ms.

2. Experimental stimuli

21 voices (phonation samples of the sustained vowel /a/) were selected from Kay Elemetrics Disordered Voice Database (Kay Elemetrics, Inc, Lincoln Park, NJ). Out of these 21 voices, ten represented distinct points along a continuum of perceived vocal roughness and had been used in prior experiments on the perception of vocal roughness (Eddins and Shrivastav, 2010). The remaining 11 voices spanned a wide range of perceived vocal breathiness and had also been used in previous perceptual experiments on vocal breathiness (Shrivastav and Sapienza, 2003; Patel *et al.*, 2012). Each stimulus was

edited to obtain a 500-ms segment over which the waveform had a relatively stable gross temporal envelope based on visual inspection. Since dysphonic voices are often unstable, choosing a short and stable stimulus helps to minimize the acoustic variability within each stimulus. These stimuli were originally recorded at a sampling rate of 50 000 Hz, but were downsampled to 24 414 Hz to match the permissible sampling rate of the hardware used for perceptual experiments. Stimuli were shaped with a 20-ms cosine-squared window to avoid any onset and offset clicks during stimulus playback.

Sample stimuli are shown in Fig. 2, with the corresponding waveform (left), magnitude spectrum (middle), and autocorrelation function (right;).[3] As noted by Yost *et al.* (1996) and Patterson *et al.* (1996), the pitch strength of a stimulus is related to the height of the first peak in the autocorrelation. Thus, the tone anchor (row 1) should have the strongest pitch strength, followed by the intermediate pitch strengths of the IRN sample (row 2, 8 dB attenuation) and the two voice samples (rows 3 and 4), while the noise anchor (row 5) should have the weakest pitch strength. The voice samples in rows 3 and 4 have quasi-periodic waveforms, reflected in low frequency portion of the spectra that highlight the harmonic nature of the sustained English /a/ vowel (as in the word /hot/). This waveform periodicity, in turn, is related to the quasi-periodic vibration of the vocal folds subsequently filtered by the vocal tract. The quasi-periodic vowel sounds have the most robust pitch sensation of any speech sound (imagine uttering the speech sound /a/ while varying from low to high on a musical scale), and reflect the pitch properties of the phoneme itself as well as the pitch properties characteristic of an individual voice.

It is important to note that pitch strength estimates using anchor stimuli frequently involve comparisons of stimuli within a trial that differ in subjective sound quality. For example, the stimuli used by Shofner and Selas (2002) within a single listening trial consisted of white noise, IRN, and a harmonic complex with equal-amplitude components below 10 000 Hz. These three stimuli differ substantially in their sound quality, acoustic characteristics, and pitch strength. Likewise, the white noise, voice tokens, and pure tone stimulus used in the main experiment here differ in sound quality. The voice tokens studied here, consisting of a sustained /a/ sound, are similar to the characteristic buzzy quality of a harmonic complex and elicit a pitch sensation that varies across tokens.

E. Preliminary evaluation: Pitch matching

While it is intuitive that the voiced sound of a sustained vowel has a distinct pitch, the specific perceived pitch of the 21 voice samples used in this study was unknown. Using a simple pitch matching task, five listeners judged the perceived pitch of

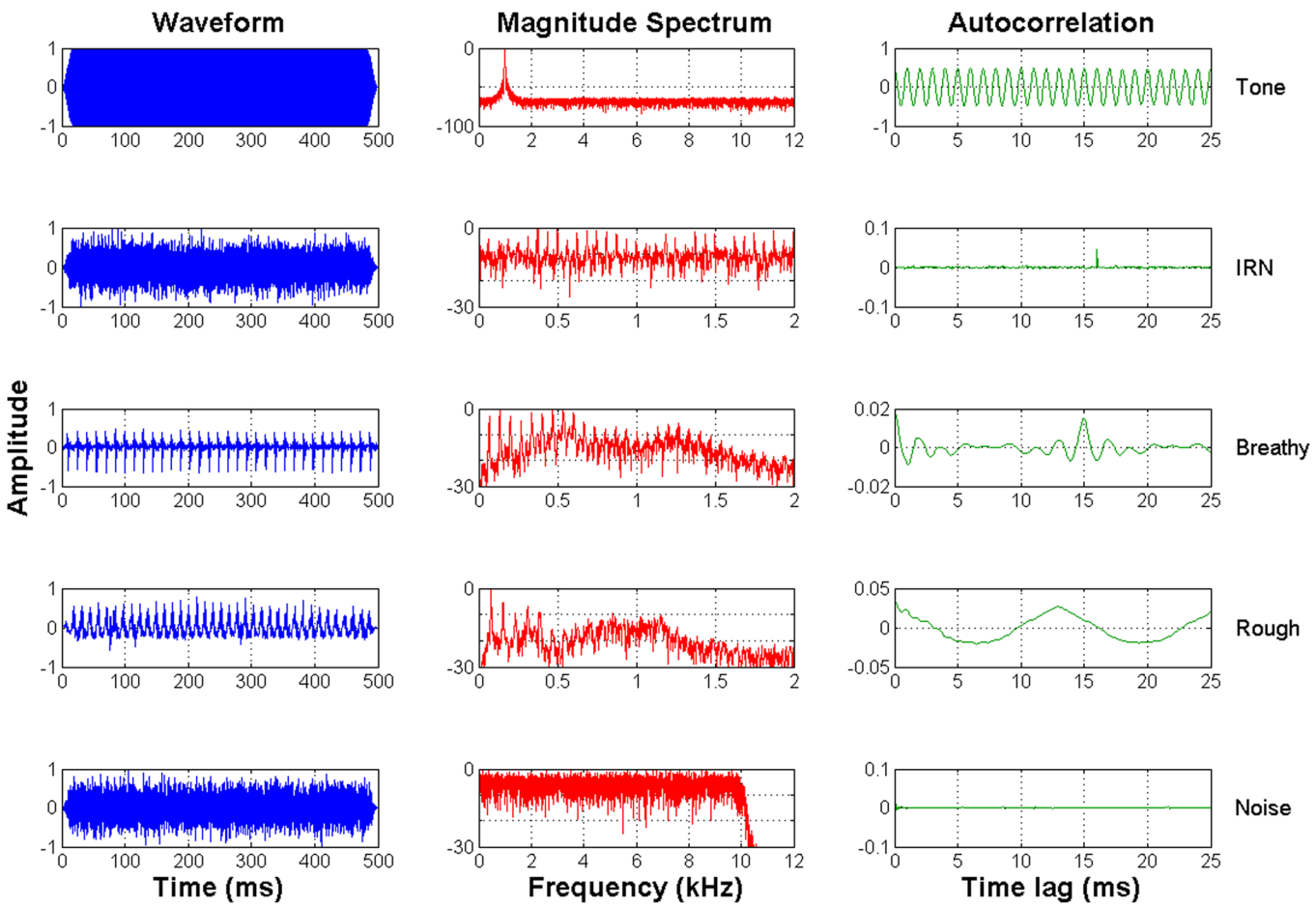

FIG. 2. (Color online) Sample waveforms (left column), magnitude spectra (middle column), and autocorrelation functions (right column) for representative stimuli as labeled to the right of rows 1–5. Note that the *y*-axis ranges for the magnitude spectrum and autocorrelation functions (middle and right column) are stimulus dependent.

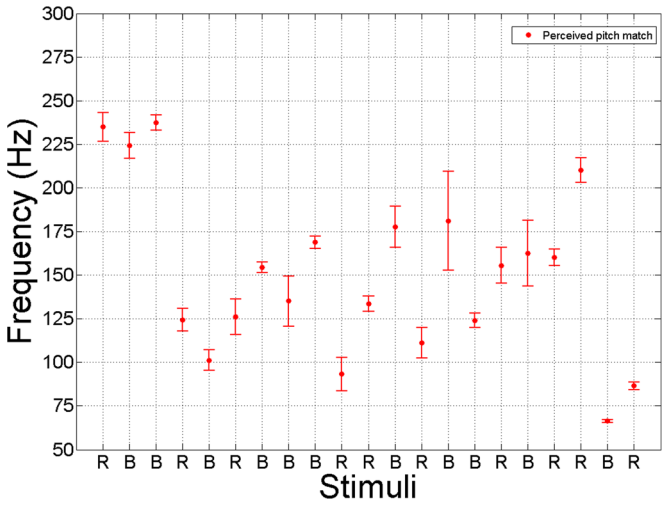

FIG. 3. (Color online) Pitch matching judgments for the 21 speech standard stimuli used in the main experiment. Each of the 21 standards are shown on the abscissa, labeled as being from either the breathy (B) or rough (R) continuum, and ordered from low to high perceived pitch strength (shown in Fig. 5). The reference sound was a complex tone with equal amplitude harmonics and variable fundamental frequency. Symbols indicate the perceived pitch match averaged across five listeners and bars indicate standard error.

the 21 voice samples described above. On each trial, listeners heard two stimuli: a reference sound consisting of one of 21 voice samples, and a comparison sound consisting of an equal-amplitude complex tone with harmonics ranging from a (variable) fundamental frequency to 4000 Hz. Listeners were asked to increase or decrease the frequency of the comparison tone such that the perceived pitch of the tone approximated the perceived pitch of the reference vowel sound. The frequency of the matching stimulus was varied according to the subject response in steps of 50, 20, and 2 Hz. The initial frequency of the comparison tone was randomly chosen over the range of 50 to 500 Hz and the final pitch match value was based on the average of three separate pitch matches. The reference stimuli were presented in random order across participants. Five participants (three male, two female) volunteered for this evaluation. None were part of the main experiment. One was the second author, and listeners ranged in age from 23 to 46 years. The results of this preliminary experiment are shown in Fig. 3 with voice sample from 1 to 21 on the abscissa and frequency (Hz) on the ordinate. The labels B and R on the abscissa indicate that the voice samples are from the breathy (B) or rough (R) continuum and correspond to the axis in Fig. 5 below. The symbols indicate the perceived pitch match averaged across five listeners. It is clear that listeners were able to assign a pitch to each voice sample and that perceived pitch varied substantially among the 21 voice samples. The average pitch matches were strongly correlated with estimates of the fundamental frequency of the individual speech tokens ($r = 0.97$).

III. RESULTS

A. Pitch strength judgments

1. Training stimuli

Previous experiments have shown that variations in the attenuation (gain) parameter in the IRN stimulus generation procedure produce stimuli that systematically vary in perceived pitch strength (Leek and Summers, 2001; Shofner and Selas, 2002; Yost *et al.*, 1978; 1979; 1996; Yost, 1997). Therefore, the training session included a set of five IRN stimuli differing in terms of the degree of attenuation used on each iteration of the stimulus generation procedure. The results of this training session are shown in the upper panel of Fig. 4 (squares) and indicate that these listeners judge pitch strength to decrease systematically along the continuum of IRN attenuation values. Importantly, these data show that the listeners grasp the concept of pitch strength and scale pitch strength as a function of IRN attenuation factor in the expected manner. These results are similar in form to those of Shofner and Selas (2002) who also explored the perceived pitch strength of IRN as a function of attenuation value despite the fact that they used a different method for generating IRN. While there are several algorithms for computing IRN (e.g., Shofner and Selas, 2002; Yost *et al.*, 1996), the algorithm used here (adopted from Yost *et al.*, 1996) has not been used to explore the effect of the attenuation (gain) parameter *per se*. Data from Shofner and Selas (2002) are plotted as the circles in the upper panel of Fig. 4. Differences in the functions from the two studies may be attributed to the use of different stimulus generation methods. The lower panel of Fig. 4 shows the height of the first peak in the autocorrelation function computed for the stimuli used in the present study as well as those used in the study of Shofner and Selas (2002). Differences in the two functions relating autocorrelation to attenuation mirror those relating perceived pitch strength to attenuation, supporting the notion that perceived pitch strength is related to the height of the first autocorrelation peak and supporting the conclusion that differences in the present data and those of Shofner and Selas (2002) are due to stimulus generation methods.

2. Experimental stimuli

The pitch strength judgments for the experimental stimuli are shown in Fig. 5, with pitch strength judgment on the ordinate and the test stimulus indicated on the abscissa, ordered from low to high perceived pitch strength. The box plots represent the mean across the ten listeners, the 25th and 75th percentile, plus/minus one standard deviation. Judgments for the 21 voices covered a broad continuum of pitch strength, spanning the range between the broadband noise and pure tone anchor stimuli. Stimuli from the breathy subset ranged in pitch strength from 17.2 to 86.0, with an average [standard deviation (SD)] score of 57.7 (23.8). Stimuli from the roughness subset were perceived to have pitch strength ranging from 20.8 to 84.1, with an average (standard deviation) of 62.0 (20.1).

To determine the inter-judge reliability, pair-wise Pearson's correlation coefficients were computed for the average pitch strength ratings for each stimulus across all pairs of listeners. These were then averaged to obtain the mean inter-judge reliability and was found to be 0.87 (SD = 0.06). Similarly, intra-judge reliability was estimated by calculating the average Pearson's correlation coefficient (r) between the ten

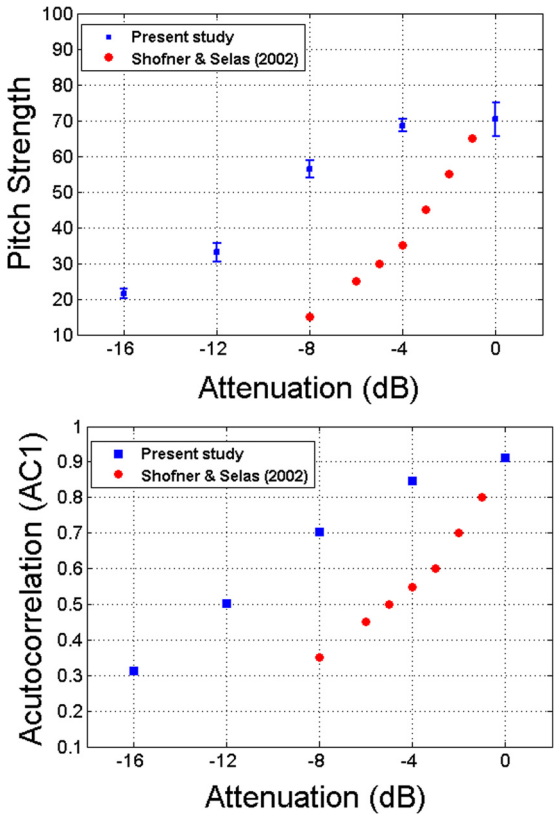

FIG. 4. (Color online) Results of the training task using IRN stimuli and the anchors from the main experiment. In the upper panel, perceived pitch strength is on the ordinate and the IRN attenuation parameter is on the abscissa. Symbols indicate perceived pitch strength using the anchored magnitude estimation task for the current study (squares, bars show standard error) and data from Shofner and Selas (2002, circles). In the lower panel, the value of the first peak in the autocorrelation function is plotted against the attenuation parameter using the same symbols as the upper panel.

repetitions of each stimulus. The mean intra-judge reliability was observed to be 0.80 (SD = 0.09). The high correlations within and across listeners showed that listeners were able to judge pitch strength in a similar and reliable manner.

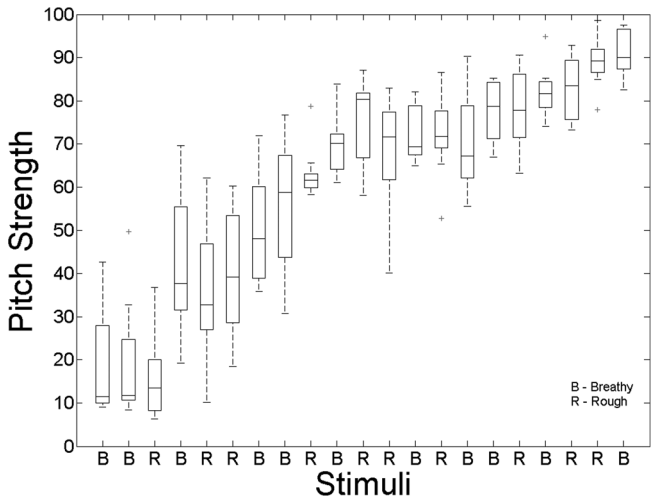

FIG. 5. Perceived pitch strength for 21 speech tokens. Each of the 21 standards are shown on the abscissa, labeled as being from either the breathy (B) or rough (R) continuum, and ordered from low to high perceived pitch strength. Box-plots based on results from the ten listeners shown the mean, first and last quartile, and standard deviation.

B. Comparison of pitch strength judgments to previous voice quality judgments

Having estimated the pitch and pitch strength of the 21 voice samples above, it is instructive to compare those estimates to previous estimates of the perceived vocal breathiness and vocal roughness for the same stimuli. Recall that the 21 voice samples used here were chosen because they vary along a continuum of normal to dysphonic voice and because the same samples have been used previously in experiments investigating perceptual judgments of breathy voice quality (Patel *et al.*, 2010) and rough voice quality (Eddins and Shrivastav, 2010). For the 11 stimuli from the breathy continuum, we have used both magnitude estimation and matching tasks to evaluate perceived breathiness. For the ten stimuli from the roughness continuum, we have used rating scale and matching tasks to evaluate perceived roughness. For simplicity, data obtained from the psychophysical matching tasks for both voice quality attributes will be compared to the current pitch strength estimates. Details of the data collection procedures for the matching tasks are described in Patel *et al.* (2010) (breathiness) and Eddins and Shrivastav (2010) (roughness). Briefly, listeners evaluated the degree of breathiness or roughness by comparing the voice samples to a synthetic comparison stimulus. In each case, the synthetic stimulus consisted of a sawtooth waveform that was lowpass filtered (151 Hz; −7 dB/octave) and mixed with similarly filtered speech-shaped noise. For estimating vocal breathiness, listeners adjusted the level of the noise with respect to the sawtooth wave [i.e., the signal-to-noise ratio or (SNR)] to match the perceived breathiness of the comparison stimulus to that of the standard stimulus (i.e., voice sample). The corresponding SNR (in dB) served as the index of breathiness (e.g., analogous to the loudness of 1000 Hz tone as an index of loudness). Likewise, roughness was evaluated by comparing the standard voice stimulus to a synthetic comparison stimulus comprised of a sawtooth + noise carrier that was amplitude modulated with a exponential (power of 4) sine function (25 Hz). Listeners varied the depth of amplitude modulation of the comparison stimulus such that the perceived roughness matched that of the standard voice sample. Thus, modulation depth (measured in dB) served as an index of the vocal roughness. These matching procedures were preferred over other measures such as rating scales or visual analog scales because the matching procedure provided ratio-level data that was relatively unbiased by context and because the index provided a physical metric useful in subsequent modeling the perception of dysphonic voices (Patel *et al.*, 2010).

Figure 6 shows perceived pitch strength judgments from the present experiment as a function of perceived breathiness (SNR in dB) for the 11 voice samples taken from (Patel *et al.*, 2010). Here, high breathiness matching thresholds correspond to less perceived breathiness (see labels on abscissa of Fig. 6). The correlation of 0.989 ($p < 0.001$) between vocal breathiness matching thresholds and mean pitch strength judgments indicates that pitch strength is inversely related to the magnitude of perceived vocal breathiness. In other words, stimuli with greater breathiness are perceived to have lower pitch strength.

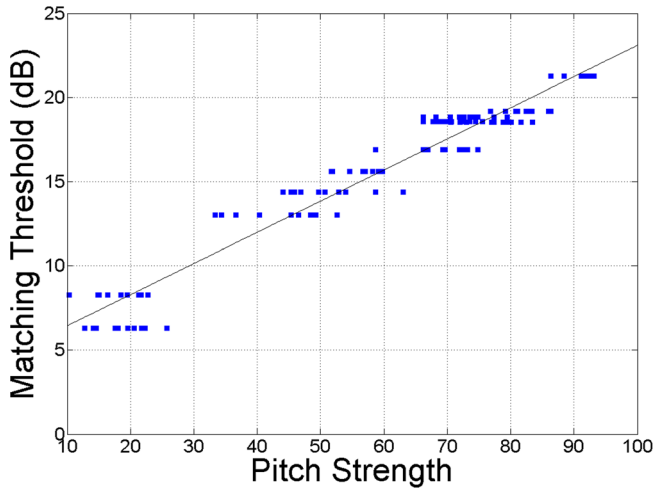

FIG. 6. (Color online) Perceived pitch strength from Fig. 5 plotted against perceived breathiness obtained using a psychophysical matching task for the 11 stimuli along a continuum of vocal breathiness. Individual pitch strength judgments for the 10 listeners (symbols) cluster around 11 points on the ordinate corresponding to the average breathiness judgments reported by Patel *et al.* (2012) for the same 11 stimuli. Matching thresholds are reported in units of signal-to-noise ratio in dB (see text for details) where values near 0 dB correspond to high perceived breathiness and values near 25 dB correspond to low perceived breathiness.

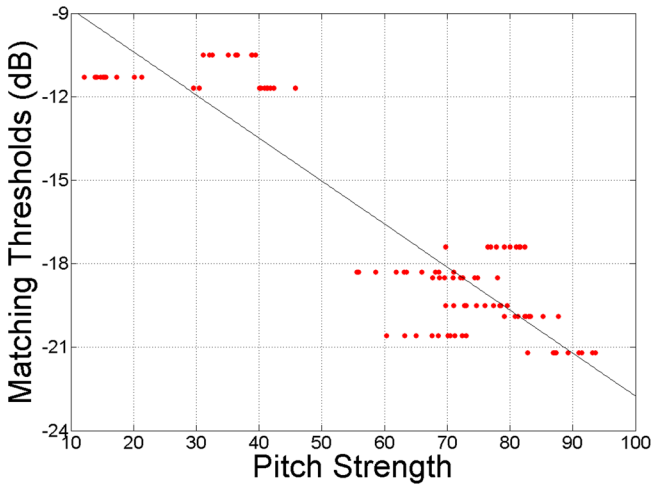

FIG. 7. (Color online) Perceived pitch strength from Fig. 5 plotted as a function of perceived roughness obtained using a psychophysical matching task for the ten stimuli along a continuum of vocal roughness. Individual pitch strength judgments for the ten listeners (symbols) cluster around ten points on the ordinate corresponding to the average roughness judgments reported by Eddins and Shrivastav (2010) for the same ten stimuli. Matching thresholds are reported in units of amplitude modulation depth in dB (see text for details) where –9 dB corresponds to high perceived roughness and – 24 dB corresponds to low perceived roughness.

Similarly, the relationship between pitch strength and roughness matching thresholds is shown in Fig. 7 for the ten stimuli that varied along a continuum of normal to disordered vocal roughness. The correlation between mean perceived pitch strength and perceived roughness matching thresholds was again strong (Pearson's $r = -0.898$; $p < 0.005$). It is evident from the figure that pitch strength is inversely related to the magnitude of roughness. In other words, stimuli with low pitch strength are perceived to have greater roughness. A linear function best described the relationship between pitch strength scores and roughness matching thresholds, accounting for 80.7% of the variance in roughness matching thresholds. However, the high numerical values may be slightly inflated for the roughness stimuli as data in the figure reveal two distinct clusters of high and low pitch strength. Indeed, when the three stimuli in the lower right portion of the graph are omitted, a linear function accounts for only 13.9% of the variance, though analyses based on only seven points should be interpreted with caution as well. Importantly, in choosing these ten rough voice samples, no attempt was made to control for covariation of roughness and breathiness (a frequent occurrence with dysphonic voices), so it is unknown whether or not the observed relationship reflects a relationship between pitch strength and vocal roughness *per se* or simple reflects a variation in breathiness with roughness in these samples.

IV. DISCUSSION

The present study was designed to explore the potential relationship between pitch strength and voice quality and was motivated by prior research to understand the perception of breathiness in vowels (Shrivastav and Sapienza, 2003; Cummings *et al.*, 2008; Shrivastav *et al.*, 2007; Shrivastav *et al.*, 2011). The current results demonstrate that listeners are capable of scaling the pitch strength of voices that vary in voice quality and that pitch strength judgments vary systematically with variations in voice quality.

Yost (1996) and Patterson *et al.* (1996) highlighted the relationship between the height of the first peak in the autocorrelation function (AC1) and pitch strength judgments. If it is assumed that such a relationship holds for non-speech as well as speech sounds, then demonstration of a similar relationship between pitch strength judgments for voiced speech and autocorrelation would lend support to the assumption here that listeners were indeed judging pitch strength and not some other perceptual attribute. Indeed pitch strength judgments were proportional to AC1, with a correlation of $r = 0.83$. Thus, similar to pitch strength measures for non-speech (IRN) stimuli, demonstration of the relationship between perceived pitch strength of voiced speech tokens and AC1 supports the notion that listeners are in fact judging pitch strength. This, combined with the high inter- and intra-judge reliability observed here provides considerable validation of the current measurement technique for use with voiced speech stimuli.

The preliminary pitch matching experiment demonstrated that listeners are quite good at matching the perceived pitch of voiced speech tokens to a complex tone of variable fundamental frequency. While it is possible that pitch strength judgments were influenced by the perceived pitch of the voice tokens as well, the correlation between pitch strength judgments and pitch match estimates was rather weak, with $r = 0.36$. This indicates that pitch, per se, was not the primary cue that listeners were using in the pitch strength task itself. A prominent acoustic feature of voiced speech is the fundamental frequency estimated from the voice token, which is related to vocal fold anatomy and physiology and gives rise to the harmonic structure of voiced speech. So it is of interest to determine the relationship between perceived pitch as estimated from the

supplemental pitch matching task described above and fundamental frequency estimates. In this case, fundamental frequency was obtained from the TF32 algorithm which is based in part on autocorrelation computations (Milenkovic, 1987). The correlation between perceived pitch as estimated from the supplemental pitch matching task described above and fundamental frequency estimates was $r = 0.97$ based on data from the five observers who completed that task. Thus, fundamental frequency was strongly related to perceived pitch but weakly related to perceived pitch strength.

Listener judgments of dysphonic stimuli showed that these stimuli exhibit a wide range of pitch strength values. While stimuli judged to have relatively normal voice quality were perceived to have high pitch strength, those with more severe breathiness or roughness were rated to have low pitch strength. There was a strong inverse relationship between pitch strength and severity of breathiness as well as a high correlation between perceived roughness and pitch strength. It is important to note that breathiness and roughness often co-occur in dysphonic voices. Since none of the stimuli tested in the current experiment were judged for both breathiness and roughness, it is difficult to ascertain whether pitch strength is correlated with breathiness per set or with both the breathy and rough voice qualities. Additional experiments to establish the relationships between breathiness and roughness are essential. Nevertheless, the wide range of pitch strength observed for dysphonic voices and the high correlation with breathiness and roughness scores indicate that inclusion of pitch strength in computational models of voice quality may improve the accuracy of their predictions of perceptual judgments.

Recent work has attempted to predict judgments of vocal breathiness using computational models that incorporate aspects of auditory processing (e.g., Shrivastav, 2003; Shrivastav and Sapienza, 2003; Shrivastav and Camacho, 2010; Shrivastav *et al.*, 2011). If the output of a computational model can accurately predict perceptual judgments, then the likelihood of both understanding the relevant perceptual processes and development of objective voice quality metrics will be increased. The models of Shrivastav *et al.* (2011) have been based on the assumption that voiced speech stimuli have both periodic (harmonic) and aperiodic (noise) elements. Accordingly, they used variants of the partial loudness model of Moore *et al.* (1997) where the partial loudness (PL) is associated with the harmonic energy of the vowel that is masked by the aperiodic components of the same voice. The noise loudness (NL) is the loudness elicited by the aperiodic components in the voice (for more details, see Shrivastav and Sapienza, 2003). The NL and PL measures computed from the loudness model are correlated with perceptual judgments of breathiness. Specifically, perceived breathiness is inversely related to PL and proportional to NL (Shrivastav, 2003; Shrivastav and Sapienza, 2003; Shrivastav and Camacho, 2010; Shrivastav *et al.*, 2011). The ratio of noise loudness to partial loudness (referred to as "η") was used as the primary predictor of perceived breathiness. The model predictions were least accurate for stimuli judged to be either very low or very high in breathiness and the model required separate parameters for male and female voices.

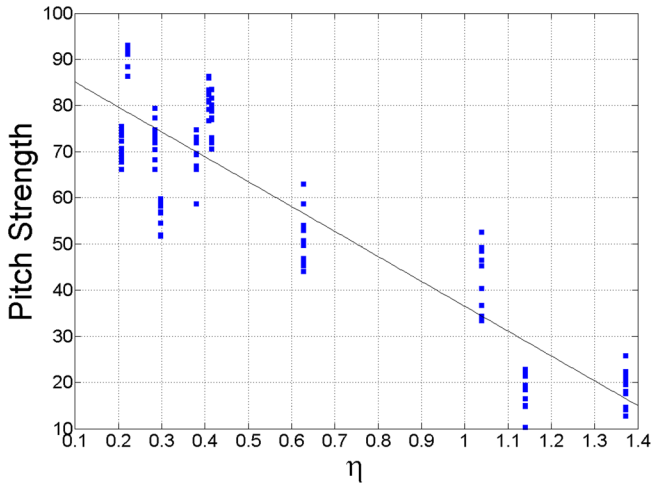

FIG. 8. (Color online) Perceived pitch strength from Fig. 5 plotted as a function predicted breathiness using the modified partial loudness model of Shrivastav *et al.* (2011). The ratio of noise loudness to partial loudness, η = NL/PL is shown on the abscissa. Symbols represent individual pitch strength estimates for the 11 stimuli from the breathy continuum.

While not evaluated here, one possibility is that if pitch strength is related to breathiness judgments, then the addition of a pitch strength parameter to such a model may improve the model predictions. The scatter plot in Fig. 8 shows the values of η computed from the model of Shrivastav *et al.* (2011) for the 11 voices from the breathy continuum plotted against individual pitch strength estimates for the same stimuli. Results show that the current mean pitch strength judgments have a high negative correlation with η ($r = -0.89$, $p < 0.001$). This reflects a moderate positive correlation with partial loudness ($r = 0.62$; $p = 0.020$) and a negative correlation with noise loudness ($r = -0.89$; $p < 0.001$). Based on these results, it is possible that inclusion of a pitch-strength estimator in the model may simplify and improve the accuracy of the model predictions of perceived breathiness.

The natural speech stimuli used here, from 21 different talkers, varied in pitch, loudness, and spectral shape. As such, they also varied in timbre. The use of a single phoneme, /a/, limited timbre differences to some extent, however, no additional attempt was made to normalize timbre. As such, the current pitch strength judgments could have been influenced by timbre differences. The use of filtered or synthetic speech would allow one to potentially control timbre differences, and such an experiment should be carried out in the future.

V. CONCLUSIONS

The voices of speakers with dysphonia vary in terms of their pitch strength. Vowels judged to have the most severe dysphonia are also judged to have to lowest pitch strength. Both breathiness and roughness were found to show a high correlation with pitch strength. Future work is required to determine if the correlation between pitch strength and perceived vocal roughness is related to roughness per se or simply the co-occurrence of breathiness in some rough voices. These findings suggest that inclusion of pitch strength in computational models of voice quality may help improve the accuracy of these models.

ACKNOWLEDGMENTS

This research was supported by a grant from NIH (Grant No. R01 DC009029). The authors wish to thank Mark Skowronski and three anonymous reviewers for their helpful comments on the manuscript.

[1]Recruitment of participants was carried out without respect to gender. Because the student make-up within the University of Florida Department of Speech, Language, and Hearing Sciences is approximately 95% female, by chance, all participants were female.

[2]While some studies have used a harmonic complex tone as the anchor corresponding to high pitch strength, here a 1000 Hz pure tone was chosen as the upper anchor so that the pitch itself was well outside the pitch range of the voice tokens used in the main experiment.

[3]The auto-correlation peak at a lag of 1 fundamental period (AC1) was measured for all stimuli by first computing an unbiased auto-correlation sequence for each stimulus using the entire 0.5-s stimulus then searching for the peak value of the sequence at lags corresponding to the range $[0.6 \times f0, 1.6 \times f0]$ where f0 was an estimate of fundamental period measured by hand. Each auto-correlation sequence was scaled such that the auto-correlation value at zero lag was 1.

ANSI. (**1994**). ANSI S1.1-1994, American National Standard Acoustical Terminology (American National Standards Institute, New York), p. 34.

ANSI. (**2004**). ANSI S3.21-2004, Methods for Manual Pure-Tone Threshold Audiometry (American National Standards Institute, New York).

ASHA. (**2002**). Consensus Auditory-Perceptual Evaluation of Voice (CAPEV) (American Speech-Language and Hearing Association, Rockville, MD).

Colton, R. H., and Casper, J. K. (**1996**). "Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment, 2nd ed. (Williams & Wilkins, Baltimore), p. 17.

Cummings, S., Patel, S., Eddins, D. A., and Shrivastav, R. (**2008**). Refining a Reference for Perceptual Measurement of Breathy Voice Quality (Voice Foundation Symposium, Philadelphia, PA).

Eddins, D. A., and Shrivastav, R. (**2010**). "Psychometric functions for rough voice quality," J. Acoust. Soc. Am. **127**(3), **2021**.

Fastl, H. (**1988**). "Pitch and pitch strength of peaked ripple noise," in *Basic Issues in Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Wit (Academic, London), pp. 370–379.

Fastl, H., and Stoll, G. (**1979**). "Scaling of pitch strength," Hear. Res. **1**, 293–301.

Houtsma, A. J. M. (**1997**). "Pitch and timbre: Definition, meaning, and use," J. New Music Res. **26**, 104–115.

Kay Elemetrics, Corp. (**1994**). Disordered Voice Database. Model 4337, 03 ed.

Kempster, G. B., Gerratt, B. R., Verdolini-Abbott, K., Barkmeier-Kraemer, J., and Hillman, R. E. (**2009**). "Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol," Am. J. Speech Lang. Pathol. **18**, 124–132.

Kent, R. (**1996**). "Hearing and believing: some limits to the auditory perceptual assessment of speech and voice disorders," Am. J. Speech Lang. Path. **5**(3), 7–23.

Klatt, D. H., and Klatt, L. C. (**1990**) "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc.Am. **87**(2), 820–857.

Kreiman, J., and Gerratt, B. R. (**2000**). "Measuring voice quality," in *Voice Quality Measurement*, edited by R. D. Kent and M. J. Bell (Singular, San Diego), pp. 73–101.

Leek, M. R., and Summers, V. (**2001**). "Pitch Strength and pitch dominance of iterated rippled noises in hearing-impaired listeners," J. Acoust. Soc. Am. **09**, 2944–2954.

Milenkovic, P. (**1987**). "Least mean square measures of voice perturbation," J. Speech Lang. Hear. Res. **30**, 529–538.

Milenkovic, P. (**2001**). TF32 [Computer software], Madison, WI.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**(4), 224–239.

Patel, S., Shrivastav, R., and Eddins, D. A. (**2010**). "Perceptual distances of breathy voice quality: A comparison of psychophysical methods," J. Voice. **24**(2), 168–177.

Patel, S., Shrivastav, R., and Eddins, D. A. (**2012**). "Developing a single comparison stimulus for matching breathy voice quality," J. Speech Lang. Hear. Res. Available: http:\\jslhr.asha.org.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). "The relative strength of tone and noise components of iterated rippled noise," J. Acoust. Soc. Am. **100**, 3286–3294.

Shofner, W. P., and Selas, G. (**2002**). "Pitch strength and Stevens' power law," Percept. Psychophys. **64**, 437–450.

Shrivastav, R. (**2003**). "The use of an auditory model in predicting perceptual ratings of breathy voice quality," J. Acoust. Soc. Am. **17**(4), 502–512.

Shrivastav, R., and Camacho, A. (**2010**). "A computational model to predict changes in breathiness resulting from variations in aspiration noise level," J. Voice **24**(4), 395–405.

Shrivastav, R., Camacho, A., Patel, S. A., and Eddins, D. A. (**2011**). "A model for prediction of breathiness in vowels," J. Acoust. Soc. Am. **125**, 1605–1615.

Shrivastav, R., Eddins, D. A., and Patel, S. (**2007**). "Developing a reference for perceptual measurement of breathy voice quality," Voice Foundation Symposium, Philadelphia, PA.

Shrivastav, R., and Sapienza, C. (**2003**). "Objective measures of breathy voice quality obtained using an auditory model," J. Acoust. Soc. Am. **114**(4), 2217–2224.

Yost, W. A. (**1982**). "The dominance region and ripple noise pitch: A test of the peripheral weighting model," J. Acoust. Soc. Am. **72**, 416–425.

Yost, W. A. (**1996**). "Pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **100**, 3329–3335.

Yost, W. A. (**1997**). "Pitch strength of iterated rippled noise when the pitch is ambiguous," J. Acoust. Soc. Am. **101**, 1644–1648.

Yost, W. A., and Hill, R. (**1978**). "Strength of pitches associated with ripple noise," J. Acoust. Soc. Am. **64**, 485–492.

Yost, W. A., and Hill, R. (**1979**). "Models of the pitch and pitch strength of ripple noise," J. Acoust. Soc. Am. **66**, 400–410.

Yost, W. A., Hill, R., and Perez-Falcon, T. (**1978**). "Pitch and pitch discrimination of broadband signals with rippled power spectra," J. Acoust. Soc. Am. **63**, 1166–1173.

Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066–1078.

Yost, W. A., Sheft, S., and Patterson, R. D. (**1994**). "Iterated rippled noise: testing theories of complex pitch," J. Acoust. Soc. Am. **95**, 2966.

Zwicker, E., and Fastl, H. (**1990**). "Pitch and pitch strength," in *Psychoacoustics: Facts and Models* (Springer-Verlag, New York), pp. 103–132.